**Göğüs Kanseri Sınıflandırması**

**Classification Nedir?**

Classification (sınıflandırma), makine öğrenmesi algoritmalarının veriyi çeşitli kategorilere ayırmak için kullanılmasıdır. Spam filtreleme, el yazısı tanıma, hastalık teşhisi gibi alanlarda kullanılır. Algoritmalar test verileri ve sonuçları arasında ilişkiler kurup sonradan aldığı, daha önceden görmediği verileri bu ilişkilere dayanarak sınıflandırması mantığıyla çalışır.

**Veriyi Araştırma:**

Bu veri, Dr. William Wolberg'in 2,5 seneyi biraz aşkın klinik çalışmaları sonucu bir araya gelmiştir. Senede 2-3 defa güncellenen bu veri 15 Temmuz 1992'de çalışmalarda kullanılması için bağışlanmış olup güncellenme geçmişi şu şekildedir:

* **Grup 1:** 367 veri (Ocak 1989)
* **Grup 2:** 70 veri (Ekim 1989)
* **Grup 3:** 31 veri (Şubat 1990)
* **Grup 4:** 17 veri (Nisan 1990)
* **Grup 5:** 48 veri (Ağustos 1990)
* **Grup 6:** 49 veri (Güncellendi: Ocak 1991)
* **Grup 7:** 31 veri (Temmuz 1991)
* **Grup 8:** 86 veri (Kasım 1991)

Dr. Wolberg'in kendi tahlilleri üzerine dayandığı için güvenilir bir veri olarak değerlendirebiliriz. 2010'dan 2019'a kadar alıntılanmış olması da genel olarak güvenilir olduğunun bir göstergesi olarak sayılabilir.

Alıntılandığı makaleleri incelediğimizde veri biliminde kullanılan çeşitli algoritma ve metotların test ve incelemelerinde kullanıldığını görüyoruz. Bu türden çalışmalar için tercih edilmiş olmasında verinin nispeten pek de büyük olmaması ve genel olarak güvenilir bulunması etkili olmuş olabilir. Bunlar da bu veriyi bizim de güvenilir bulmamız için yeterli sebepler.

Özellikleri genel olarak incelersek:

* **Clump\_thickness:** Kümenin kalınlığı.
* **Uniformity\_of\_cell\_size:** Hücre boyutlarının düzenliliği.
* **Uniformity\_of\_cell\_shape:** Hücre şekillerinin düzenliliği.
* **Marginal\_adhesion:** Kanser hücrelerinin çevre dokulara bağlılık durumu.
* **Single\_epithelial\_cell\_size:** Epitel hücrelerinin boyutu.
* **Bare\_nuclei:** Hücre çekirdeklerinin durumu.
* **Bland\_chromatin:** Homojen veya düz chromatin dağılımı.
* **Normal\_nucleoli:** Nükleollerin morfolojik özelliği.
* **Mitoses:** Hücre bölünme süreci.
* **Class:** Kanserin türünü belirleyen, sınıf değeri.

Özelliklerin değer aralığı:

* **Clump\_thickness:** 1 – 10
* **Uniformity\_of\_cell\_size:** 1 – 10
* **Uniformity\_of\_cell\_shape:** 1 – 1
* **Marginal\_adhesion:** 1 – 1
* **Single\_epithelial\_cell\_size:** 1 – 10
* **Bare\_nuclei:** 1 – 10
* **Bland\_chromatin:** 1 – 10
* **Normal\_nucleoli:** 1 – 10
* **Mitoses:** 1 – 10
* **Class:** Kanserin türünü belirleyen, sınıf değeri.